



## Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

# FREQUENCY DISTRIBUTIONS OBTAINED BY CERTAIN TRANSFORMATIONS OF NORMALLY DISTRIBUTED VARIATES.\*

BY H. L. RIETZ.

The problem considered in this paper was first suggested to the writer by experiments with actual frequency distributions of various measurements of objects which approximate roughly to a set of similar solids. To be concrete, we may think of the diameters, surfaces, and volumes of spheres that represent objects in nature, such as oranges on a tree or peas on a plant.

Suppose the distribution of diameters is a normal distribution given by

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}.$$

It seems natural to inquire into the nature of the distribution of the corresponding surfaces and volumes. Conversely, we should ask for a determination of the distribution of diameters if we knew surfaces or volumes were normally distributed. The same kind of problem† would arise if we knew that velocities,  $v$ , of molecules of a gas were normally distributed, and were required to investigate the distribution of energy  $\frac{1}{2}mv^2$ . These concrete illustrations are special cases of the transformation of variates of a normal distribution by replacing each variate,  $x$ , by an assigned function  $kx^n$ , where  $k$  is a positive constant and  $n$  is a positive integer or the reciprocal of a positive integer. Edgeworth‡ and Kapteyn§ have made use of transformations of the normal curve as a method of representing skew frequency distributions. Apart from the possible use for this purpose, the frequency curves arising from certain simple transformations of the variates of a normal distribution present points of special interest to which it seems that attention should be directed, particularly because of the striking differences in general appearance from normal curves—a fact that seems both interesting and important in forming a proper conception of the place of the normal curve in the representation of frequency distributions.

It is the main purpose of the present paper to exhibit certain properties

---

\* Read before the American Mathematical Society at Lincoln, Nebraska, Nov. 27, 1920.

† Edgeworth, Proc. Fifth International Congress of Mathematicians, II, p. 427.

‡ Loc. cit. and a series of papers in the Journal of the Royal Statistical Society. See vol. 61, pp. 670–700.

§ Skew Frequency Curves in Biology and Statistics, 1903.

of the frequency curves that are obtained when the variates of a normal distribution are transformed by substituting for each variate,  $x$ , the function  $kx^n$  where  $k$  is a positive constant, and where suitable restrictions will be placed on  $n$  as we proceed. The case  $n = 1$  is treated by Bruns\* and the results are simple and well known. Edgeworth called attention to the general form of the frequency curve with which we are concerned for  $n = 2$ . Furthermore, when deviations of variates from their mean value are small compared to their mean value, it is well known that the distributions of squares and cubes of variates approach normal distributions sufficiently near for certain purposes. But in certain important statistical applications, the deviations of variates from their mean value cannot be reasonably regarded as small compared to the mean value. This latter class of distributions gives special importance to our problem.

When  $k = 1$ , the problem is that of exhibiting the properties of the frequency distribution of the  $n$ th powers of a set of normally distributed variates. This case seems to include essentially the same points of interest contained in the more general problem, since the transformation  $x' = x^n$ , followed by the linear transformation  $x'' = kx'$ , produces the same result as the transformation  $x'' = kx^n$ . Hence we shall in what follows deal with the transformation  $x' = x^n$ .

To determine the frequency function obtained by the transformation, let  $x_1, x_2, \dots, x_i$  be a system of variates expressed in a unit equal to the standard deviation  $\sigma$ , and belonging to the normal distribution

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2}}, \quad \bar{x} \equiv 0,$$

so that  $P = \int_a^b y dx$  is the probability that a variate taken at random belongs to the interval  $a$  to  $b$ .

Let us replace each variate  $x_s$  by  $x_s'$ , where  $x_s' = x_s^n$ . We then make a corresponding transformation of the integral  $\int_a^b y dx$  by letting  $x' = x^n$  ( $n \neq 0$ ). Then

$$dx' = nx^{n-1}dx, \text{ except at } x = 0 \text{ when } n < 1,$$

and

$$dx = \frac{dx'}{nx'^{\frac{n-1}{n}}}, \text{ except at } x' = 0 \text{ when } n > 1.$$

---

\* Wahrscheinlichkeitsrechnung und Kollektivmasslehre, 1906, pp. 126-129.

We may therefore write

$$P = \frac{1}{\sqrt{2\pi}} \int_a^b e^{\frac{-(x-\bar{x})^2}{2}} dx = \frac{1}{n\sqrt{2\pi}} \int_{a^n}^{b^n} e^{\frac{-\left(\frac{1}{x'^{\frac{1}{n}}-\bar{x}}\right)^2}{2}} dx', \quad (1)$$

where for the present we assume  $a \equiv 0$ , and  $b \equiv 0$ . As shown below, these limitations on  $a$  and  $b$  may be removed to some extent for certain values of  $n$ .

The frequency curve of  $x'$ -variates obtained from positive  $x$ -variates is then given by

$$y' = \frac{1}{n\sqrt{2\pi} x'^{\frac{n-1}{n}}} e^{\frac{-\left(\frac{1}{x'^{\frac{1}{n}}-\bar{x}}\right)^2}{2}}. \quad (2)$$

The function (2) does not represent a normal curve when  $n \neq 1$ .

In order to determine the general character of the frequency curve given by (2), we first examine the function for maxima and minima. For this purpose, we have

$$\frac{dy'}{dx'} = \frac{1}{n\sqrt{2\pi}} e^{\frac{-\left(\frac{1}{x'^{\frac{1}{n}}-\bar{x}}\right)^2}{2}} \left[ -\frac{n-1}{n} x'^{-\frac{2n-1}{n}} - \frac{1}{n} (x'^{\frac{1}{n}} - \bar{x}) x'^{\frac{2-2n}{n}} \right].$$

The derivative changes signs at

$$x' = \frac{1}{2^n} (\bar{x} \pm \sqrt{\bar{x}^2 - 4(n-1)})^n, \quad (3)$$

when  $\bar{x}^2 > 4(n-1)$ , and at  $x' = 0$  for certain values of  $n$ .

In equation (1) we restricted  $a$  and  $b$  to be zero or positive, but when  $n$  is an odd positive integer or the reciprocal of an odd positive integer, it follows at once that (2) gives the frequency curve corresponding to negative values of  $x'$  that arise from the transformation  $x' = x^n$  when  $x$  is negative. By taking  $\bar{x}$  sufficiently large, the function (2) may be made as nearly zero as we please for negative values of  $x'$  except at points near the discontinuity at  $x' = 0$ . This discontinuity exists when  $n > 1$ . When  $n$  is an odd number or the reciprocal of an odd number, the derivative  $dy'/dx'$  changes sign at  $x' = 0$ . When  $n$  is the reciprocal of an odd positive integer, there is a minimum at  $x' = 0$ , and the value of the function is zero at this minimum.

We shall find it convenient to consider the frequency curves given by (2) under three cases according as  $n > 1$ ,  $0 < n < 1$ , or  $n < 0$ . We shall

limit our discussion to positive values of  $x$  and  $x'$  except when there is a specific statement extending the treatment to negative values.

CASE I.  $n > 1$ .

The maximal frequency corresponds to the value of  $x'$  given by taking the positive sign before the radical in (3). In the language of statistics, the abscissa of this maximal frequency is called the modal value or the mode. We shall find it convenient to use these expressions later in this paper. The curve for  $n = 3$ ,  $\bar{x} = 4$  is shown in Fig. 1 for positive values of  $x'$ .

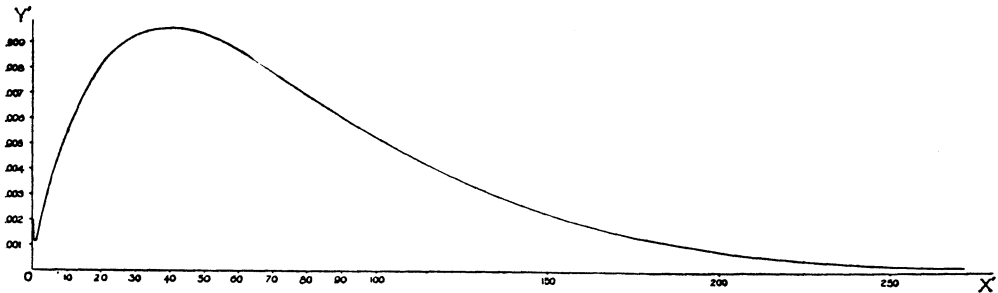


FIG. 1.  $y' = \frac{x'^{-2/3}}{3\sqrt{2}\pi} e^{\frac{-(x'^{1/3}-4)^2}{2}}$ ,  $x' > 0$ .

The skew appearance of the figure shows that the distribution is not even approximately a normal distribution. Since variates at the median of the original normal distribution must be transformed to the median of the new distribution, we may appropriately compare the value  $\bar{x}^n$  of the median of the new distribution with the modal value given by (3). When  $n > 1$ , it follows from (3) that the mode of the new distribution is less than its median. The minimum that corresponds to the value of  $x'$  given by taking the negative sign before the radical in (3) is of special interest because the existence of this minimum would probably not be expected by ordinary intuition. Thus, when  $\bar{x}^2 > 4(n-1)$ , we have a minimum between the origin and the maximum discussed above. This minimum is shown in Fig. 1 at a point near the origin for the case  $n = 3$ ,  $\bar{x} = 4$ . The descent of the curve from infinity at  $x' = 0$  to the minimum at  $x' = (2 - \sqrt{2})^3$  is so rapid and the curve is so near the Y-axis that it cannot be shown well on the scale of Fig. 1. For this reason we show the curve in the neighborhood of this minimum in Fig. 2 on an enlarged scale. The function  $y'$  ( $n$  an odd number) has real positive values when  $x'$  is negative, but the values differ very little from zero, except near  $x' = 0$ , when  $\bar{x} \cong 4$ . For the case  $\bar{x} = 4$  shown in Fig. 1, the values of  $y'$  for

negative values of  $x'$  are so near zero, except at points near the discontinuity at  $x' = 0$ , that it is impractical to exhibit the curve on the scale of Fig. 1 for negative values of  $x'$ .



FIG. 2. Portion of curve of Fig. 1 from  $x' = 0$  to  $x' = 1$  with enlarged horizontal scale.

Next we find  $d^2y'/dx'^2$ . It turns out that the points of inflection are given by the solutions of the equation

$$x'^{\frac{1}{n}-3} [x'^{\frac{4}{n}} - 2\bar{x}x'^{\frac{3}{n}} + (3n - 4 + \bar{x}^2)x'^{\frac{2}{n}} + \bar{x}(3 - 3n)x'^{\frac{1}{n}} + (n - 1)(2n - 1)] = 0. \quad (4)$$

When  $n = 3$ ,  $\bar{x} = 4$ , one point of inflection is at  $x' = 1$  as shown in Fig. 1. Furthermore, it is easily verified when  $\bar{x} = n + 1$ , that (4) has a solution  $x' = 1$ , and that there is a point of inflection at  $x' = 1$ . There is also, in general, another point of inflection on the curve to the right of the maximum.

The general appearance of the frequency curve (2) depends much on the value of  $\bar{x}$  compared to  $4(n - 1)$ . When  $\bar{x}^2 \equiv 4(n - 1)$ , the function (2) has no maximum nor minimum, but is a monotone decreasing function of  $x'$ . When  $\bar{x}^2 = 4(n - 1)$ , there is a point of inflection at  $x' = \bar{x}^n/2^n$ .

The problem when  $n = 2$  and  $\bar{x} > 2$  presents a point of special interest. Thus, if  $\bar{x}$  were assigned larger and larger values, the  $x$ -coördinate of the minimum would approach zero and that of the maximum would approach the median  $\bar{x}^2$ . This is seen from the fact that

$$\frac{1}{4}(\bar{x} - \sqrt{\bar{x}^2 - 4})^2$$

and

$$\bar{x}^2 - \frac{1}{4}(\bar{x} + \sqrt{\bar{x}^2 - 4})^2.$$

are monotone decreasing functions of  $\bar{x}$ .

An analogous result holds for the minimum when  $n > 2$ , but it does

not hold for the maximum. Thus, when  $n$  has any assigned value  $> 2$ , the  $x$ -coördinate

$$x' = \frac{1}{2^n} (\bar{x} - \sqrt{\bar{x}^2 - 4(n-1)})^n$$

of the minimum is a monotone decreasing function of  $\bar{x}$ , and approaches zero as a limit when  $\bar{x}$  is increased indefinitely, but the mode does not, in general, approach the median  $\bar{x}^n$  as a limit when  $n$  is increased. However, the ratio

$$\frac{1}{2^n \bar{x}^n} (\bar{x} + \sqrt{\bar{x}^2 - 4(n-1)})^n$$

of the mode to the median approaches the limit 1 as  $\bar{x}$  is increased indefinitely.

The rapidity of approach to the limiting values depends on the smallness of the ratio  $4(n-1)/\bar{x}^2$ . Hence, in order that the frequency curve may descend rapidly to a minimum in the neighborhood of the discontinuity at  $x' = 0$ , and in order that the mode shall be relatively near the median, it is necessary that  $4(n-1)/\bar{x}^2$  shall be small. This condition is clearly necessary in order that the new frequency curve shall have roughly the appearance of a normal curve when we neglect the part of this curve which belongs to the interval from  $x' = 0$  to the minimum.

CASE II.  $0 < n < 1$ .

In this case, make  $n = 1/m$ , where  $m > 1$ . This case thus includes the distribution obtained by taking positive integral roots of a set of variates. We shall limit our considerations to the principal real values of the functions.

The equation (2) may be written

$$y' = \frac{m x'^{m-1}}{\sqrt{2\pi}} e^{\frac{-(x'^m - \bar{x})^2}{2}}. \quad (5)$$

When  $n < 1$ , it follows from (3) that the mode is greater than the median of the new distribution. There is a minimum at  $x' = 0$  when  $m$  is an odd number  $> 1$ , and we have in this case a minimum given by the negative value of  $x'$  obtained from (3). If  $4(n-1)/\bar{x}^2$  is small, the value of the function for  $x' < 0$  is too nearly zero to distinguish the curve from the  $x$ -axis when drawn on a scale suitable for reproduction on an ordinary page. Further, if  $4(n-1)/\bar{x}^2$  is small, the curve for  $x' > 0$  may be described roughly as having the general appearance of a normal curve, but differing from the normal curve both in that it is somewhat skew, and in that  $y' = 0$  at a finite point.

CASE III.  $n < 0$ .

Let  $n = -m$ .

Then the frequency curve becomes

$$y' = \frac{1}{\frac{m+1}{mx'^m} \sqrt{2\pi}} e^{-\frac{\left(x' - \frac{1}{m-x}\right)^2}{2}}. \quad (6)$$

By giving  $y'$  the value zero when  $x' = 0$ , the curve becomes continuous at the origin.

The distribution has a modal value

$$x' = \frac{(\sqrt{\bar{x}^2 + 4(m+1)} - \bar{x})^m}{2^m(m+1)^m}.$$

This mode is less than the median  $1/\bar{x}^m$ .

From the three cases examined relative to values of  $n$ , we may now state the theorem that *the  $n$ th powers of a set of normally distributed positive variates give a distribution whose modal value is greater or less than its median according as the value of  $n$  is or is not between 0 and 1.*

The simplicity of the examination of the frequency distributions obtained from a normal distribution by the transformation  $\bar{x} = x^n$  arises from the fact that the equation

$$\frac{dy}{dx} = 0$$

has a quadratic factor in the variable  $x = x'^{\frac{1}{n}}$  for which we solved to determine maxima and minima.

The occurrence of this quadratic factor suggests the problem of finding other functions

$$x' = f(x) \quad (7)$$

which would lead to a quadratic equation in  $x$ , and in more special cases to a linear equation in  $x$ , for finding maxima and minima of the frequency distribution obtained by the transformation (7).

Assume that (7) may be solved for  $x$  giving a single-valued function

$$x = \varphi(x'). \quad (8)$$

Then the frequency curve of  $x'$ -variates is

$$y' = \frac{1}{\sqrt{2\pi}} e^{-\frac{[\varphi(x') - \bar{x}]^2}{2}} \frac{dx}{dx'},$$

and

$$\frac{dy'}{dx'} = \frac{1}{\sqrt{2\pi}} e^{-\frac{[\varphi(x') - \bar{x}]^2}{2}} \left\{ \frac{d^2x}{dx'^2} - \left( \frac{dx}{dx'} \right)^2 (x - \bar{x}) \right\} = 0 \quad (9)$$

gives the maxima and minima.

We may now seek the function that will make the equation

$$\frac{d^2x}{dx'^2} - \left(\frac{dx}{dx'}\right)^2 (x - \bar{x}) = 0 \quad (10)$$

have a quadratic factor in  $x$ .

Such a quadratic factor would exist, in general, if  $\left(\frac{dx}{dx'}\right)^2 / \frac{d^2x}{dx'^2}$  were a linear function of  $x$ , say  $cx + c_1$ .

That is,

$$cx \frac{d^2x}{dx'^2} + c_1 \frac{d^2x}{dx'^2} = \left(\frac{dx}{dx'}\right)^2. \quad (11)$$

Let  $p = dx/dx'$ . Then (11) becomes

$$cxp \frac{dp}{dx} + c_1 p \frac{dp}{dx} = p^2,$$

and apart from the trivial solution  $x = a$  constant, we have the solutions

$$x' = c_2 \left(x + \frac{c_1}{c}\right)^{1-\frac{1}{c}} = c_2 [x + c_1(1 - n)]^n, \quad (12)$$

and

$$x' = c_3 \log \left(x + \frac{c_1}{c}\right). \quad (13)$$

Thus we find that the logarithms of the variates as well as their powers are distributed in accord with frequency curves whose maxima and minima are easily obtained because of the quadratic factor in (10) when  $x' = \log x$ . The frequency distribution for the transformation  $x' = \log x$  is similar to that of the case  $0 < n < 1$  discussed above in that the mode is greater than the median.

When  $\left(\frac{dx}{dx'}\right)^2 / \frac{d^2x}{dx'^2} = c_1$ , where  $c_1$  is a constant, the equation (10)

has, in general, a linear factor, and

$$x' = c_4 e^{-\frac{x}{c_1}} + c_5. \quad (14)$$

Thus we find that a simple exponential transformation of variates leads to a linear factor in equation (10).

Another transformation that would, in general, lead to a quadratic factor in (10) is given by making

$$\frac{\frac{d^2x}{dx'^2}}{\left(\frac{dx}{dx'}\right)^2} = Ax^2 + Bx + C.$$

From this equation

$$x' = c_6 \int \frac{dx}{e^{\frac{Ax^3}{3} + \frac{Bx^2}{2} + Cx}},$$

which could hardly be regarded as a simple transformation in which we are likely to be interested unless  $A = B = 0$ , but this special case gives simply the transformation (14).

THE UNIVERSITY OF IOWA,  
IOWA CITY, IOWA.